SAMGrid Operations Policies

Adam Lyon
January 14, 2008
v1.0

This document describes the policies for operating the SAMGrid system.

General support summary:

Time	Support Level	Comments	Paging
FNAL business hours	~1 hour response	Expert shifter monitors sam-oncall mailing list and issue tracker	Expert shifter may be paged/called as necessary
Weekday off hours	Best effort	Expert shifter checks sam-oncall mailing list and IT once in evening	Expert shifter should not be called/paged before 9am or after 9pm except for data handling online emergency
Weekends and FNAL holidays	Best effort	If possible, expert shifter checks sam-oncall list and IT at least once per day	Expert shifter should not be called/paged before 9am or after 9pm except for data handling online emergency

1 SAMGrid Operations level of service

While MC and primary production are extremely important, those activities are not as critical as the storage of online data from the detector to tape. Failure to store online data is the only emergency situation for REX/Ops. Expert shifters on call may be contacted at all hours for such problems. Problems in other components, including SAMGrid, do not warrant all hours contact of experts. The level of service for SAMGrid is described here and summarized in the table above.

1.1 General practices

When an issue tracker ticket is assigned to the on call expert, the expert should post a comment to the ticket stating that he/she is investigating. The on call expert will investigate the problem, assign it to others if necessary, and eventually resolve the ticket when the problem itself is resolved. The on call expert is responsible for that ticket (regardless of other assignees) until their shift period is over. If the ticket is still unresolved at that time, it is passed to the next shifter. If the unresolved ticket is awaiting action from another assignee, then the current on call expert is responsible for

making sure that action is completed.

If the on call expert is overloaded, he or she may transfer the ticket to the secondary on call expert. That secondary shifter then has responsibility for the ticket. The secondary shifter should acknowledge ownership by adding a comment to the ticket.

Note that this system assumes that all issues appear in the Issue Tracker. If an issue arises that is not in the IT, the expert shifter should ask the issue contact person to submit to the issue tracker by sending mail to d0sam-admin@fnal.gov.

1.2 Response during Fermilab business hours

During regular Fermilab business hours (9am-5pm)[1], the expert shifter on call will monitor the sam-oncall mailing list. Issue tracker tickets assigned to the on call expert by the regular SAM shifter will be acknowledged in a reasonable time (~1 hour).

1.3 Response during weekday off hours

If possible, the on call shifter should check their mail and the Issue Tracker sometime in the evening and attempt to handle problems that have emerged. Note that most likely problems will not be resolved until the next business day, but a best effort should be put forward. Also note that most of the hardware at Fermilab supporting SAMGrid is NOT supported 24/7 by system administrators.

Since SAMGrid is not a detector critical system, the on call expert should not be called before 9am or after 9pm for SAMGrid problems.

1.4 Response during weekends and Fermilab holidays

Response during this time is best effort. The on call expert should check e-mail and the Issue Tracker at least once per day and respond to problems is possible. If the on call expert will not have internet access, he or she should make arrangements with the secondary shifter to do this work. The on call expert should not be called between the hours of 9pm to 9am on weekends and holidays for SAMGrid problems.

Note that the resolution of problems may have to wait until the next business day. Also note that most of the hardware at Fermilab supporting SAMGrid is NOT supported 24/7 by system administrators.

2 A process is used to triage and assign tickets

The process for triaging and assigning tickets flows as follows:

A user has a problem. He or she either sends mail to d0sam-admin@fnal.gov (automatically creating a new ticket) or creates a ticket with the Issue Tracker web page. All problem reports should go through the issue tracker system so that the issue can be tracked. People sending issues directly to shifters will be told to resend the problem

1 All times in this document are US Central Time.

report to the issue tracker.

The DØ offline shifter acknowledges the ticket and attempts to solve it. Either the ticket is resolved, or the offline shifter assigns the ticket to the expert on call.

The expert shifter can wait for the ticket to be assigned to them, or address it immediately when it appears in the issue tracker system. In either case, the expert shifter acknowledges that they own the ticket.

The expert shifter works the problem, assigns the ticket to others if necessary, and ultimately ensures that the ticket is resolved.

3 Deployments and changes to production systems are made by the REX/Ops group

When the transition to REX/Ops is complete, the developers should no longer access the production systems except for troubleshooting and monitoring. Development for new components and bug fixes should occur on designated development and test nodes (see below), not production. Deployments on the production system will be made by the REX/Ops group, with advice and assistance from the developers if necessary. The developers should not change any item on the production system without permission from the SAMGrid project manager.

3.1 Rationale

Given that REX/Ops is responsible for uptime of production SAMGrid, we must be in complete control of the production system. Deployments must be made by us, so we can schedule downtimes with the stake holders. Furthermore, if we make the deployments we can run our tests and pay closer attention to watch for possible problems caused by the new components.

There are potential negatives of this plan. Because REX/Ops makes the production deployments and not the developers, new features and bug fixes may not be immediately available to the stake holders. Doing these deployments makes for more work for our group, especially if many deployments are necessary to fix a problem.

4 Deployments are made to production only after testing

In the complicated distributed environment of SAMGrid, full scale production level testing is nearly impossible. There is no test farm that can mimic all of the environments and variables that the production system faces. Therefore, some failures of new code are, unfortunately, unavoidable. However, we should try to catch at least obvious problems before new code is deployed. We would like to have confidence that new code, once deployed, will not fail immediately or cause problems with other services. To this end, we will perform a limited integration test on all new code before deploying to production.

4.1 The test system

We will build a limited test system to replicate some of the production SAMGrid functionality in a mixed test-production environment. We will have a test queuing node (like samgrid.fnal.gov), a test forwarding node (right now just OSG, though we may add an LCG forwarding node in the future), and if necessary a test durable cache code (though little SAMGrid software resides on a durable cache node). The queuing and forwarding nodes are the ones with the most SAMGrid installed software, so it makes the most sense to have test versions of them. Since we have no test worker node farm, we will have to submit test jobs to production execution sites. Therefore, we will need to use the production SAM stations and data delivery infrastructure. See section 6 for more information on this test system.

This test system will be used for continuous testing of SAMGrid (see section 5 below), testing new code for deployment (this section), and if needed, loaned to the developers for their integration testing before releases (e.g. for a new version of VDT).

4.2 Procedures

When a new version of SAMGrid code is ready for release, we will deploy it first onto the test queueing node and/or the test forwarding node. We will then launch two standard test jobs, one for Monte Carlo production and the other for collider data processing. We will monitor the progress of the test jobs. If the jobs complete successfully, the new code is ready to be deployed into production. Any failures must be investigated and tests repeated until success.

4.3 Implications

A result of testing before deployments is that bug fixes and new features may not be immediately available for the end users – they will have to wait for the testing to complete. In general, such a delay for testing is good because it gives us, and the end users, more confidence in the new release. The integration testing described here should not take more than several days (e.g. less than one week).

If the production system is in an inoperable state due to a failure of the SAMGrid code, the SAMGrid project manager has the option to deploy fixes directly to production and then do the tests along with production jobs. For such a situation, the priority is to get the system working again as quickly as possible. The end users will be warned that there may be further bugs or problems that may cause problems with their jobs and necessitate more releases. If the system is in a usable state, deployments must be tested first on the test system, never deployed directly to production. Even small, seemingly simple changes must be integration tested first.

5 Continuously test the production system

We will use the test system described above for periodic submission of test jobs through the production system managed by the REX/Ops group. The test jobs should be “real”

enough to download an executable and read in at least one input file as well as creating an output file (that can be later discarded). The OSG and LCG forwarding systems should be tested in parallel. High volume native SAMGrid sites should also receive test jobs. It also would be nice to have some automated way to check the results of these jobs.

6 A test system is used for development and integration testing

There should be several nodes that parallel main components of the production system. These nodes may be used for development and testing of new components.

When there is no development (e.g. “peacetime”), the test nodes should have the same components and versions as found as in the production system. These nodes may then be used to launch test jobs to ensure the integrity of the test nodes deployment.

If the developers need the test nodes for some development, they may borrow the nodes from REX/Ops. Note that REX/Ops may require downtimes at any time for these nodes for system maintenance as per the system administrators.

When the developers are finished with their use of the nodes, they should restore the nodes to their “peacetime” state. They should NOT deploy new versions of components to be left for REX/Ops.

If new version of components are to be deployed, REX/Ops will deploy them first to the test nodes, if feasible. The nodes must be relinquished by the developers before this time. Test jobs will then be launched. If they succeed along with other testing, then the production system will be upgraded as per consultations with stake-holders.